Fed-MUnet: Multi-modal Federated Unet for Brain Tumor Segmentation

Ruojun Zhou*, Lisha Qu*, Lei Zhang*, Ziming Li†, Hongwei Yu†, Bing Luo*

*Duke Kunshan University, Kunshan, China
†Wuhan University, Wuhan, China

Abstract—Deep learning-based techniques have been widely utilized for brain tumor segmentation using both single and multi-modal Magnetic Resonance Imaging (MRI) images. Most current studies focus on centralized training due to the intrinsic challenge of data sharing across clinics. To mitigate privacy concerns, researchers have introduced Federated Learning (FL) methods to brain tumor segmentation tasks. However, currently such methods are focusing on single modal MRI, with limited study on multi-modal MRI. The challenges include complex structure, large-scale parameters, and overfitting issues of the FL based methods using multi-modal MRI. To address the above challenges, we propose a novel multi-modal FL framework for brain tumor segmentation (Fed-MUnet) that is suitable for FL training. We evaluate our approach with the BraTS2022 datasets, which are publicly available. The experimental results demonstrate that our framework achieves FL nature of distributed learning and privacy preserving. For the enhancing tumor, tumor core and whoe tumor, the mean of five major metrics were 87.5%, 90.6% and 92.2%, respectively, which were higher than SOTA methods while preserving privacy. In terms of parameters count, quantity of floating-point operations (FLOPs) and inference, Fed-MUnet is Pareto optimal compared with the state-of-the-art segmentation backbone while achieves higher performance and tackled privacy issue. Our codes are open-sourced at https://github.com/Arnold-Jun/Fed-MUnet.

Index Terms—Brain tumor segmentation, Federated Learning, Multi-modality, BraTS2022 dataset.

I. INTRODUCTION

SEGMENTING brain tumors from multi-modal Magnetic Resonance Imaging (MRI) images is clinically important for cancer diagnosis and treatment. However, the variability in tumor location, shape, and appearance complicates precise segmentation. The advent of deep learning (DL) has led to the proposal of numerous advanced technologies for brain tumor segmentation. Since MRI imaging can be multi-modal by nature if acquired using different imaging sequences, such as T1-w, T1c, T2-w, and FLAIR, an increasing amount of researchers focus on modern architecture to deal with multi-modal input, as exemplified by [1]–[4]. The great success of these multi-modal methods relies highly on large-scale and centralized datasets.

However, it is typically challenging to directly gather and exchange medical images across hospitals or organizations in real-world practice. Strict privacy protection regulations, including the Health Insurance Portability and Accountability Act (HIPAA) [5] and the General Data Protection Regulation (GDPR) [6], make it difficult to share medical imaging data across hospitals. For instance, the European Union (EU) GDPR clearly stipulates that the transfer of personal data is limited to the EU or countries/organizations. To address privacy issues, McMahan [7] first proposed a Federated Learning (FL) framework. In FL, the raw data is kept on the local device and not uploaded to the central server or elsewhere. Only intermediate results, such as model parameters or gradient updates, are encrypted and transmitted, which cannot be directly translated to the original data, thus protecting data privacy. Recent works [8]–[10] implemented FL in medical tasks, aiming to address the privacy issue in brain tumor segmentation. However, the methods in [8], [9] perform tumor segmentation using the single MRI as input due to complexity of multi-modal images. The method in [10] directly uses U-Net [11] for segmentation task on multi-modal images without further exploration of modern multimodal models suitable for federated learning.

Consequently, there is a lack of study on the application of FL to multi-modal MRI brain tumor segmentation tasks. There are two major challenges in introducing FL to multi-modal brain tumor segmentation model:

• First, some multi-modal models have a two-stage structure, which greatly reduces the efficiency of FL training. For instance, recent work [4] proposed pre-training a variational autoencoder (VAE) and used the pre-trained VAE in the segmentation backbone. When combined with FL, this type of two-stage model increases the communication overhead and increases the risk of training interruption due to network anomalies or equipment failures. From this perspective, a one-stage end-to-end architecture is preferred in the FL setting.

• Second, the existing multi-modal segmentation models have a large number of parameters for dealing with multi-modal input, which are not conducive to FL training. In FL, overfitting is a common problem. The large number of parameters in multi-modality models such as DeepLabv3+ [2] exacerbates overfitting.

To address the above challenges, in this paper, we propose a novel transformer-based FL framework (Fed-MUnet) for multi-modal brain tumor segmentation. Our model follows client-server architecture, and clients train the local dataset using a newly proposed segmentation backbone, Multi-modal U-net (M-Unet). M-Unet leverages the advantages of U-Net, such as efficient structure and fewer parameters, while having the capability of multi-modal feature fusion. To mitigate the problem of overfitting in FL, we design a lightweight Cross Modality Module (CMM) that helps the model achieve SOTA

segmentation accuracy without significantly increasing the number of model parameters in the FL paradigm. CMM takes advantage of the powerful ability of transformer to capture global information, which effectively multi-modal image features. Some advanced deep learning models have been proposed in multi-modal brain tumor segmentation, such as introVAE [4] and Deeplabv3+ [2]. However, most of these models do not fit the FL framework. We summarize the key contributions of work as follows:

- We develop an end-to-end segmentation model **M-Unet** incorporating a **CMM** that effectively addresses the low training efficiency issue and alleviates overfitting in FL. To the best of our knowledge, we are the first to integrate FL with multi-modality in brain tumor segmentation. With the lightweight model, we improve the performance of segmentation accuracy without increasing the complexity and number of parameters.
- We propose a novel FL framework **(Fed-MUnet)** for multi-modal brain tumor segmentation, where differential privacy is implemented to maintain the confidentiality of patient data. This approach balances high segmentation performance with privacy protection.
- The evaluation results on BraTS2022 show that our framework achieved comparable accuracy to centralized learning methods: for the enhancing tumor, tumor core and whole tumor, the mean of five major metrics were **(87.5%)**, **(90.6%)** and **(92.2%)**, respectively, which were higher than SOTA methods.

II. PRELIMINARY

In this section, we first present basic information about brain tumor segmentation. In the second part, we detail the FL setting for our model deployment.

A. Background

Brain tumor segmentation is a very important part in clinical medicine. In order to better diagnose brain tumors, MRI data sets are widely used in the diagnosis and treatment of brain tumor diseases in hospitals. MRI has the characteristics of multi-parameter and multi-sequence imaging, which can reflect different information of the tumor area and provide a rich data basis for brain tumor segmentation. Our research aims to develop a privacy-preserving high-precision multi-modal FL framework to process MRI datasets for clinical diagnosis.

B. Federated Learning Setting

Let $\mathcal{D}$ denotes the MRI dataset and each dataset X_K for client K is random sampled from $\mathcal{D}$. Let w_K represents the proportion of the entire dataset $\mathcal{D}$ for X_K assigned to the K-th client, $w = \{(w_1, w_2, \ldots, w_N) \mid w_i \geq 0, \sum_{i=1}^{N} w_i = 1, \text{for } i = 1, 2, ..., N\}$. The vector w is a concrete sample drawn from dirichlet distribution, that is, $w = (w_1, w_2, \ldots, w_N) \sim$
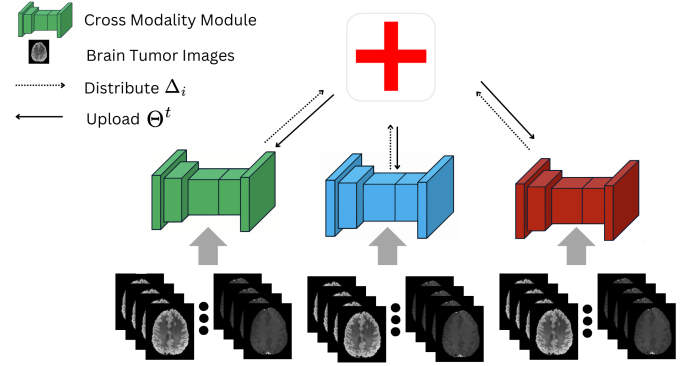

Fig. 1: Decentralized Training Paradigm for Brain Tumor Segmentation.

$Dir(w \mid \alpha).$[1] For total modality $\mathcal{M}$, local dataset$X_K = \{X_K^1, X_K^2, ..., X_K^{\mathcal{M}}\}$.

Fig. 1 illustrates the decentralized training paradigm of our model, which adheres to a typical client-server structure. Each hospital performs forward propagation using local datasets X_K and local model θ_K. After transferring the computed parameter differential Δ_K back to the central server for global aggregation, the updated model Θ^t is then sent to the local client for next global training epoch.

III. METHOD

Aiming to achieve privacy preservation, we adopt differential privacy (DP) as a powerful privacy protection mechanism for secure transmission and aggregation of model parameters. In order to tackle the challenge that multi-modal models are difficult to deploy in the FL, we propose a novel segmentation model Multi-modal U-net (M-Unet), which is primarily based on transformer and U-net.

In this section, we illustrate FL algorithm our local segmentation model M-Unet in detail. High performance achieved by our model Fed-MUnet can be attributed to M-Unet. In the first part, we elaborate on the specific algorithms for updating and transmitting model parameters. Then, we explain the original intention of the structure design. In the third part, we describe the operation mechanism and design reason of CMM.

A. Federated Learning with Differential Privacy

In our work, we utilize DP-FedAvg [12] to achieve privacy protection. The specific procedures are detailed in Algorithm 1 and Algorithm 2.

1) Client Side: On the client side, initialized/aggregated parameters Θ^t are transferred from server for local training. Notably, the clipping operation $Clip(\theta, \Theta^t)$ ensures that there are no extreme values or outliers that could significantly skew the results. Additionally, clipping restricts the maximum sensitivity of the model's output with respect to a single data

---

[1]$Dir(w \mid \alpha) = \frac{1}{B(\alpha)} \prod_{i=1}^{N} w_i^{\alpha_i - 1}$, where $B(\alpha)$ is the multivariate Beta function and hyperparameter α controls the variance of dirichlet distribution.

Algorithm 1 Client Side

1: **Input:** Global model: Θ^t, where t is the global epoch; loss function $l(\theta)$; hyperparameters: learning rate η; local epochs e; batch size: B; local model: θ; Parameters upload: $\Delta_K = \theta - \Theta^t$.
2: Initialize local model $\theta \leftarrow \Theta_0$
3: **for** each local epoch $j = 1, 2, \ldots, E$ **do**
4:     Split local data into different batches $\mathcal{B}_k$
5:     **for** each batch $b \in \mathcal{B}_k$ **do**
6:         Perform gradient descent: $\theta \leftarrow \theta - \eta \nabla l(\theta; b)$
7:         Clip model parameters: $\theta \leftarrow \Theta^t + FlatClip(\theta - \Theta^t)$ ▷ Eq.(1)
8:     **end for**
9: **end for**

---

point, thereby allowing less noise to be added while still maintaining the desired level of privacy:

$$FlatClip(\Delta_K) = \Delta_K \cdot \min\left(1, \frac{S}{\|\Delta_K\|}\right), \quad (1)$$

where S is the prior upper bound of $\|\Delta_K\|$.

Algorithm 2 Server Side

1: **Input:** Learning rate η_t, gradient clipping bound C; noise parameter σ; batch size for each round B; number of training examples generated by client k n_K; per-user example cap $\hat{w}$.
2: Randomly initialize model parameter Θ^0
3: Define client weight: w_k for client k as $w_k = \min(\frac{n_k}{\hat{w}}, 1)$
4: Set $W = \sum_k w_k$
5: **for** each round of global epoch $t = 1, 2, \ldots, T$ **do**
6:     Select a sample group of clients $\mathcal{C}^t$ for the current iteration with probability q
7:     **for** each client $k \in \mathcal{C}^t$ **do**
8:         Execute local training: $\Delta_K = \theta - \Theta^{t-1}$
9:     **end for**
10:    Aggregate client updates: $\Delta^t = Agg(q, W, \Delta_K)$ ▷ Eq. (2)
11:    Clip operation: $\Delta^t = Clip(\Delta^t, C)$ ▷ Eq. (3)
12:    Update global model parameter:
       $\Theta_t \leftarrow Update(\Theta_{t-1}, \Delta^t, \sigma)$ ▷ Eq. (4)
13: **end for**

Notably, clients results will be randomly sampled with probability q for global updating to further preserve privacy. For efficiency, in each global epoch, $q \times 10^2$ percent of clients will perform local training.

2) Server Side: On the server side, the parameter differential collected from all clients is aggregated via :

$$Agg(q, W, \Delta_K) = \frac{\sum_{k \in C^t} w_k \Delta_K}{qW}. \quad (2)$$

To prevent training instability caused by gradient explosion, we also adopt Clip operation on server side:

$$Clip(C, \Delta^t) = \Delta^t / \max(1, \frac{\|\Delta^t\|}{C}). \quad (3)$$

Finally, Gaussian noise and Δ^t are used to update the global model parameters. The variance σ is a hyperparameter to manually control the strength of privacy protection:

$$Update(\Theta_{t-1}, \Delta^t, \sigma) = \Theta_{t-1} + \Delta^t + N(0, I\sigma^2). \quad (4)$$

B. The Segmentation Backbone

We naturally adopt U-net due to its outstanding performance in medical segmentation tasks. References [11] and [13] demonstrate several advantages of U-net, including its symmetric encoder-decoder architecture, skip connections and properties suitable for few-shot learning. As shown in Fig. 2, multi-modal images are first fed into a convolutional neural network (CNN) to extract different levels of features. Due to the locality and weight sharing characteristics of convolution operation, CNN has high training efficiency when dealing with large-scale image data. To fit into the FL setting, we chose CNN as the feature extractor for downsampling instead of Transformer to reduce parameters and model complexity. Apart from the symmetric end-to-end structure, we also propose a CMM for processing multi-modal medical features. The attention-based CMM aims to integrate multi-modal features obtained from CNN for enhanced feature upsampling.

C. Cross Modality Module

CMM is mainly based on Transformer. Through the self-attention mechanism, Transformer model can directly calculate the dependency between any two locations, so as to effectively capture long-distance dependencies. Traditional CNN models, however, can only capture local information when dealing with multi-channel image. Therefore, we use Transformer-like structure in CMM to better integrate multi-modal features. As shown on the right side of Fig. 2, we start with features of four modalities, each of size $C \times H \times W$. To obtain cross-modality feature matrix, we perform the following steps:

1) **Stack the Images:** First, we stack the features $x^j_{stage_i}$ along the channel dimension to form a single tensor of size $4C_i \times H_i \times W_i$. Position embedding is added to the stacked features to help the model understand the sequential relationship between elements and patches.
2) **Flatten the Tensor:** Next, we flatten the tensor into a 2D matrix. Specifically, we reshape the tensor from $(4C_i \times H_i \times W_i)$ to $4C_i \times (H_i \cdot W_i)$.
3) **CMM Output:** The resulting matrix $\hat{x}_{stage_i}$ is used as the input to the transformer. In the last phrase of CMM, a $Conv2d$ is used to map the fused features $\hat{x}_{stage_i}$ to align the dimension of upsampling matrix $\in \mathbb{R}^{4C_i \times \frac{H_i}{2} \times \frac{W_i}{2}}$.

The loss function of our method contains two parts. The Dice loss is defined by:

$$\mathcal{L}_{Dice} = -\frac{2 \sum_{x \in \Omega} F(x) \cdot Y(x)}{\sum_{x \in \Omega} F(x)^2 + \sum_{x \in \Omega} Y(x)^2}. \quad (5)$$

Dice loss focuses on the global segmentation effect. To make the model focus on the classification accuracy of each pixel, we added pixel-wise cross entropy loss:

$$\mathcal{L}_{CE} = -\frac{\sum_{x \in \Omega} [Y(x) \log(F(x) + (1 - Y(x) \log(1 - F(x))]}{\sum_{x \in \Omega} 1}, \quad (6)$$

where the x is the pixel of image area Ω, $F(x)$ is the model output and $Y(x)$ is the ground truth.

IV. EXPERIMENTS

We train and test our model using the public dataset Brats2022. We compare our method with other methods on six metrics: Dice score, Jaccard index, Hausdorff distance, Sensitivity, Precision and Specificity. We also valuate
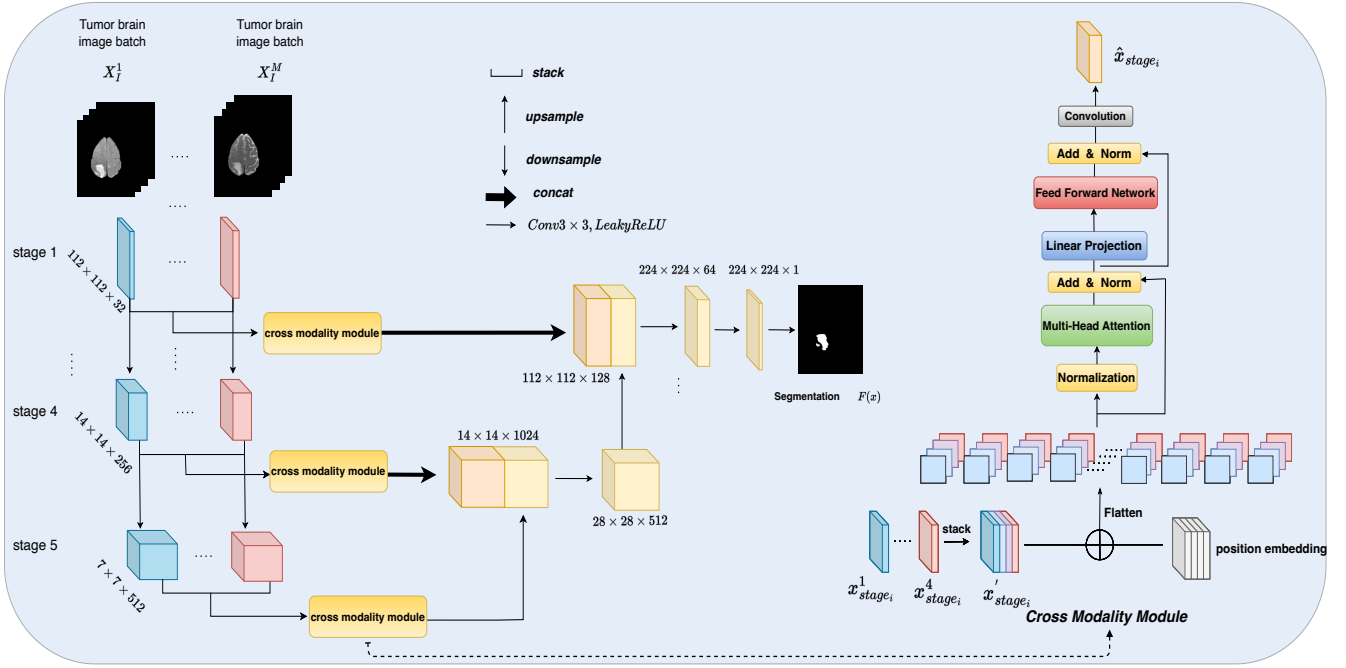
Fig. 2: The framework of our segmentation backbone M-Unet. The right part is the structure of CMM. The overall framework follows the encoder-decoder architecture of Unet and a Transformer-like module CMM is used for multi-modal feature integration. For concatenation, we deploy a convolution operation to align the dimension of upsampling feature matrix

the complexity of our model compared with other high-performance models.

A. Experimental Settings

1) Datasets: The BraTS22-GLI challenge dataset[2] is used, including multiparametric MRI (mpMRI) scans of brain tumors from routine clinical acquisitions at different centers as test, validation, and training data. The Brats dataset is a public medical image dataset provided by multiple medical centers for the research and development of brain tumor segmentation algorithms. The training and validation data are the same as those used in BraTS21. Expert neuroradiologists identify the genuine annotations of the tumor subregions. All BraTS mpMRI scans are provided as NIfTI files (.nii.gz).

A total of xyz patients are included in this study. Each patient contains MRI Images of four modalities, each with a dimension of $240 \times 240 \times 155$. The four modalities, as shown in Fig. 3 have different properties and different roles in clinical diagnosis: T1 imaging for anatomical structure; T1c for tumor enhancement by injecting contrast agent into the blood of the subject before imaging. T2 imaging for higher lesion visibility; FLAIR is a T2 hydrographic image (suppressing the high signal of the cerebrospinal fluid), which is brighter when the water content is large and can determine the peritumoral edema area.

2) Implementation details: The parameter α in Dirichlet distribution controls the sharpness of the distribution. If $\alpha < 1$, the distribution will be highly unbalanced. However, the dataset is not sufficient to set $\alpha < 1$ since some of the

[2]Since the test data of the BraTS challenge is not accessible to the public, we mix the accessible training and validation sets and randomly split all samples into 70%, 20% and 10% for training, validation and testing, respectively.

local dataset X_K may be extremely small if $\alpha < 1$, which makes it hard for the local model to learn. Therefore, in the experimental phase, we set $\alpha = 2$ to make the dataset distribution slightly more evenly.

In our implementation, all methods use the same training settings with mainstream experiments. The optimizer is SGD with a momentum of 0.95. The learning rate is 10^{-2} and the batch size is 32. In federated learning setting, we train for 30 global rounds to ensure that the model converges steadily and that the local update epoch is set to 3. In local rounds, each client K has a 50% probability (q=0.5) of being selected to perform a local epoch. Our model is trained on an RTX 4090 GPU and implemented using PyTorch.

3) Evaluation metrics: Six metrics are used to assess the performance of the model in different aspects. Specifically, the Dice score, the harmonic mean of precision and recall, is used to assess the degree of overlap between segmentation results and ground truth labels. The Jaccard index calculates the ratio of the intersection size to the union size between the segmentation result and the ground truth label. The Hausdorff distance measures the greatest separation between two edge contours, providing a more sophisticated assessment of the segmented edges' accuracy. Sensitivity is defined as the percentage of true positive cases (TP) overall true tumor regions (TP+FN). Precision measures the percentage of true positive cases (TP) over all regions predicted to be tumors (TP+FP). Specificity estimates the percentage of true negative cases (TN) among all non-tumor regions (TN+FP). Moreover, we evaluate the mean and standard deviation of the six metrics on enhancing tumor (ET), tumor core (TC), and whole tumor (WT) respectively.

TABLE I: Evaluation Results of Our Method Compared with Other SOTA Frameworks Our proposed model Fed-MUnet is decentralized and other benchmarks are centralized.

Method	Metrics	Quantification					
		Dice score(%)	*Jaccard index(%)*	*Hausdorff distance*	*Sensitivity(%)*	*Precision(%)*	*Specificity(%)*
Atten. U-Net [16]	ET	0.814±0.192	0.720±0.207	14.498±10.252	0.845±0.198	0.826±0.178	0.992±0.003
	TC	0.836±0.192	0.758±0.226	10.252±16.500	0.853±0.216	0.859±0.175	0.994±0.004
	WT	0.889±0.103	0.813±0.133	12.511±19.155	0.903±0.125	0.890±0.105	0.997±0.003
nnU-Net [17]	ET	0.829±0.167	0.751±0.254	3.582±6.597	0.819±0.255	0.875±0.140	0.994±0.005
	TC	0.875±0.136	0.813±0.247	5.219±10.210	0.856±0.243	0.917±0.122	0.996±0.006
	WT	0.919±0.060	0.858±0.113	5.974±13.016	0.899±0.113	0.939±0.062	0.996±0.003
UNETR [18]	ET	0.834±0.185	0.748±0.206	6.915±13.715	0.842±0.199	0.862±0.168	0.992±0.004
	TC	0.847±0.206	0.776±0.228	9.609±15.458	0.846±0.214	0.890±0.166	0.995±0.004
	WT	0.902±0.099	0.834±0.132	10.454±18.364	0.896±0.121	0.918±0.097	0.997±0.005
Fed-MUnet **with** $Noise \sim N(0, 10^{-2})$	**ET**	**0.840±0.171**	**0.750±0.184**	**5.134±10.191**	**0.824±0.191**	**0.880±0.150**	**0.999±0.002**
	TC	**0.878±0.144**	**0.805±0.178**	**10.347±17.100**	**0.875±0.154**	**0.908±0.129**	**0.998±0.005**
	WT	**0.905±0.094**	**0.837±0.128**	**11.383±17.110**	**0.895±0.121**	**0.925±0.075**	**0.995±0.006**
nnFormer [19]	ET	0.842±0.169	0.756±0.195	5.429±10.275	0.848±0.185	0.867±0.157	0.992±0.002
	TC	0.859±0.186	0.789±0.215	10.089±17.450	0.870±0.188	0.883±0.167	0.995±0.002
	WT	0.909±0.828	0.843±0.118	9.427±16.778	0.906±0.110	0.922±0.795	0.996±0.003
MultiFormer [20]	ET	0.843±0.181	0.760±0.204	5.060±11.903	0.839±0.202	0.882±0.153	0.994±0.002
	TC	0.856±0.196	0.786±0.223	7.383±13.025	0.848±0.210	0.908±0.149	0.995±0.003
	WT	0.911±0.087	0.846±0.120	8.820±16.547	0.904±0.115	0.927±0.078	0.997±0.005
TransBTS [3]	ET	0.850±0.177	0.770±0.200	4.330±9.708	0.844±0.201	0.892±0.139	0.994±0.003
	TC	0.873±0.189	0.810±0.214	6.043±11.971	0.859±0.205	0.921±0.146	0.996±0.005
	WT	0.913±0.081	0.849±0.115	9.096±16.448	0.913±0.104	0.921±0.079	0.996±0.004
MultiCNN [21]	ET	0.852±0.176	0.772±0.199	3.789±9.510	0.847±0.197	0.882±0.153	0.994±0.003
	TC	0.873±0.191	0.811±0.216	5.122±10.265	0.861±0.202	0.916±0.124	0.994±0.005
	WT	0.915±0.082	0.852±0.115	7.420±13.502	0.902±0.110	0.937±0.066	0.997±0.007
TuningUNet [1]	ET	0.857±0.167	0.777±0.192	3.519±6.246	0.860±0.181	0.883±0.154	0.994±0.002
	TC	0.875±0.179	0.811±0.208	5.831±9.994	0.869±0.191	0.912±0.143	0.996±0.003
	WT	0.919±0.073	0.859±0.110	5.920±11.229	0.907±0.107	0.942±0.057	0.997±0.003
Fed-MUnet **with** $Noise \sim N(0, 10^{-5})$	**ET**	**0.860±0.164**	**0.778±0.176**	**3.870±9.777**	**0.859±0.181**	**0.882±0.144**	**0.998±0.002**
	TC	**0.894±0.141**	**0.830±0.173**	**5.348±10.514**	**0.895±0.144**	**0.914±0.130**	**0.998±0.006**
	WT	**0.916±0.083**	**0.857±0.118**	**7.361±13.085**	**0.904±0.116**	**0.938±0.060**	**0.996±0.004**
Fed-MUnet **with** **no noise**	**ET**	**0.862±0.157**	**0.781±0.174**	**3.370±5.781**	**0.871±0.175**	**0.876±0.136**	**0.998±0.002**
	TC	**0.897±0.140**	**0.834±0.171**	**4.961±9.658**	**0.892±0.146**	**0.923±0.117**	**0.998±0.004**
	WT	**0.921±0.087**	**0.861±0.120**	**6.716±12.820**	**0.908±0.116**	**0.945±0.061**	**0.997±0.005**

T1c
T1
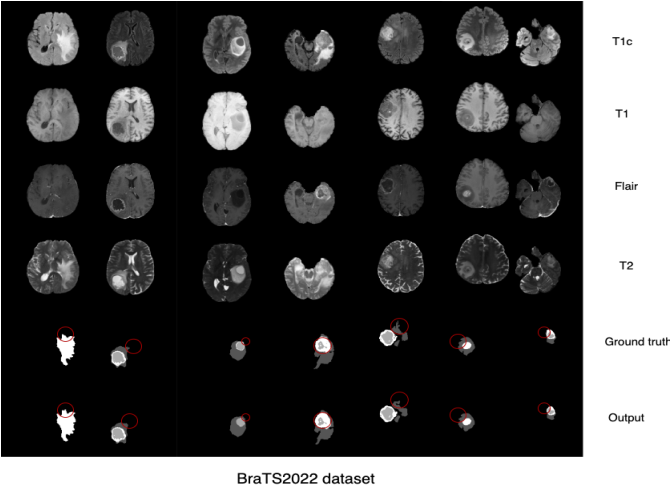Flair
T2
Ground truth
Output

BraTS2022 dataset

Fig. 3: Examples of Segmentation Results and Ground Truth. The area marked by the red circle is the difference between the true label and the inference result of the model with $\sigma = 10^{-5}$.

B. *Experimental Results*

Fig 3 displays some segmentation results of our proposed model with $\sigma = 10^{-2}$, $\sigma = 10^{-5}$ and no noise, and other centralized model have the same experimental setting.

As shown in **Table I**, our model, Fed-MUnet with $\sigma = 10^{-5}$, demonstrates substantial performance while ensuring data privacy. For the Dice Score, Fed-MUnet with noise achieves commendable values of 0.860 for Enhancing Tumor (ET), 0.894 for Tumor Core (TC), and 0.916 for Whole Tumor (WT) indicating a high degree of overlap with the ground truth. In the Jaccard Index, the model achieves results of 0.778 for ET, 0.830 for TC, and 0.853 for WT, further proving its segmentation accuracy. The Hausdorff Distance metrics demonstrates competitive performance with values of 3.870 for ET, 5.348 for TC, and 7.361 for WT. Sensitivity results, which are crucial for identifying true positive rates, are robust at 0.859 for ET, 0.895 for TC, and 0.904 for WT. Precision, reflecting the model's ability to avoid false positives, is also high, with scores of 0.882 for ET, 0.914 for TC, and 0.919 for WT. Finally, the model achieves near-perfect specificity across all tumor regions, underscoring its reliability in distinguishing non-tumor areas. Although Fed-MUnet without noise showed slightly superior results across most metrics, the marginal performance difference compared to Fed-MUnet with noise highlights the latter's robustness in protecting data privacy without significantly compromising efficacy.

Furthermore, we evaluate the complexity of our proposed framework based on the total number of parameters, the quantity of floating-point operations (FLOPs), and the

average inference time. As shown in **Table II**, our model achieves a balanced performance across these metrics while significantly improves segmentation accuracy. Specifically, Fed-MUnet contains 71.61 million parameters, which is moderate compared to other models such as MultiCNN with 177.31 million parameters and UNETR with 90.86 million parameters. The computational complexity of Fed-MUnet, as indicated by FLOPs, is 13.98 Giga, which is lower than models like nnU-Net (59.53 Giga) and MultiCNN (91.01 Giga), reflecting its efficiency in computation. Furthermore, Fed-MUnet's inference time stands at 0.186 seconds, demonstrating its competitive speed relative to other models such as UNETR (0.208 seconds) and MultiCNN (0.239 seconds). Notably, our model is the Pareto optimal except for TransBTS and MultiFormer, but achieves higher performance and tackled privacy issue.

These results demo onstrate that Fed-MUnet achieves an optimal balance between parameter count, computational demand, and inference time, making it a suitable modal for high-precision brain tumor segmentation in a federated learning framework.

TABLE II: Evaluation of Model Complexity

Method	Metrics		
	Params(M)	*FLOPs(G)*	*Inference time(s)*
Atten.U-Net [16]	34.88	31.26	0.135
nnU-Net [17]	64.14	59.53	0.159
UNETR [18]	90.86	12.86	0.208
nnFormer [19]	89.29	12.41	0.198
MultiFormer [20]	89.97	12.72	0.122
TransBTS [3]	24.23	12.84	0.117
MultiCNN [21]	177.31	91.01	0.239
TuningUNet [1]	59.12	27.65	0.188
Fed-MUnet	**71.61**	**13.98**	**0.186**

V. CONCLUSION

In this paper, we proposed Fed-MUnet for segmenting brain tumors from multi-modal MRI images. The proposed segmentation backbone M-Unet could be easily deployed in FL setting, addressing the privacy issue in medical big data. We conducted experiments using the publicly available BraTS2022 dataset to demonstrate the effectiveness of our algorithm. Our framework Fed-MUnet achieved the highest performance on brain tumor segmentation and preserved data privacy with few parameters and low model complexity.

In future work, we will investigate more sophisticated multi-modal fusion techniques to better secure data privacy without adversely affecting segmentation accuracy. In addition, the modal heterogeneity of MRI datasets mainly refers to the data difference and diversity caused by different imaging parameters, sequence types, equipment manufacturers and other factors in the MRI imaging process. We will explore advanced FL algorithm to modal heterogeneity in MRI.

REFERENCES

[1] M. Futrega, M. Marcinkiewicz, and P. Ribalta, "Tuning U-Net for brain tumor segmentation," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, S. Bakas et al., Eds. Cham, Switzerland: Springer, 2023, pp. 162–173.

[2] Choudhury, Ahana Roy, Rami Vanguri, Sachin R. Jambawalikar and Piyush Kumar.(2018). "Segmentation of Brain Tumors Using DeepLabv3+." in *BrainLes@MICCAI*.

[3] W. Wang, C. Chen, M. Ding, H. Yu, S. Zha, and J. Li, "TransBTS: Multimodal brain tumor segmentation using transformer," in *Proc. 24th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2021, pp. 109–119.

[4] H. Liu, Z. Ni, D. Nie, D. Shen, J. Wang, and Z. Tang, "Multimodal brain tumor segmentation boosted by monomodal normal brain images," in*IEEE Transactions on Image Processing*, vol. 33, pp. 1199–1210, Jan. 2024, doi: 10.1109/tip.2024.3359815.

[5] U.S. Department of Health and Human Services, "Health Insurance Portability and Accountability Act of 1996 (HIPAA)," Public Law 104-191, Aug. 1996. Available: https://www.hhs.gov/hipaa

[6] European Parliament and Council of the European Union, "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)," *Official Journal of the European Union*, L119, pp. 1-88, May 2016.

[7] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Y. Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," *International Conference on Artificial Intelligence and Statistics*, pp. 1273–1282, Apr. 2017.

[8] Y. Zhao, Q. Liu, X. Liu, and K. He, "Medical Federated Model with Mixture of Personalized and Sharing Components," in *Proceedings of the IEEE*, vol. 10, no. 5, pp. 1-10, 2022.

[9] Li, W., Milletarì, F., Xu, D., Rieke, N., Hancox, J., Zhu, W., Baust, M., Cheng, Y., Ourselin, S., Cardoso, M. J., & Feng, A. (2019). Privacy-Preserving Federated Brain Tumour Segmentation. In Suk, H.I., Liu, M., Yan, P., Lian, C. (Eds.) Machine Learning in Medical Imaging. MLMI 2019. Lecture Notes in Computer Science, vol 11861. Springer, Cham.

[10] Ullah, Faizan, Muhammad Nadeem, Mohammad Abrar, Farhan Amin, Abdu Salam, and Salabat Khan, "Enhancing Brain Tumor Segmentation Accuracy through Scalable Federated Learning with Advanced Data Privacy and Security Measures," Mathematics, vol. 11, no. 19, p. 4189, Oct. 2023.

[11] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Munich, Germany, 2015, pp. 234-241.

[12] R. Geyer, T. Klein, and M. Nabi, "Differentially Private Federated Learning: A Client Level Perspective," *arXiv:1712.07557*, 2017.

[13] Drozdzal, M., Vorontsov, E., Chartrand, G., Kadoury, S., & Pal, C. (2016). "The Importance of Skip Connections in Biomedical Image Segmentation". In *Deep Learning and Data Labelling for Medical Applications* (pp. 179-187). Springer, Cham.

[14] H. Tzu-Ming Harry Hsu and M. Brown.(2019). "Measuring the Effects of Non-Identical Data Distribution for Federated Visual Classification," arXiv, Sep. [Online]. Available: https://arxiv.org/abs/1909.06335

[15] S. Guo, X. Yang, J. Feng, Y. Ding, W. Wang, Y. Feng, and Q. Liao, "FedGR: Federated Learning with Gravitation Regulation for Double Imbalance Distribution," in *Database Systems for Advanced Applications - 28th International Conference, DASFAA 2023, Tianjin, China*, April 17-20, 2023, pp. 1-16. Springer, 2023.

[16] J. Schlemper et al., "Attention gated networks: Learning to lever- age salient regions in medical images," Med. Image Anal., vol. 53, pp. 197–207, Apr. 2019.

[17] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation," Nature Methods, vol. 18, no. 2, pp. 203–211, Feb. 2021.

[18] A. Hatamizadeh et al., "UNETR: Transformers for 3D medical image segmentation," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2022, pp. 574–584.

[19] H.-Y. Zhou, J. Guo, Y. Zhang, L. Yu, L. Wang, and Y. Yu,(2021), "NnFormer: Interleaved transformer for volumetric segmentation," ,*arXiv:2109.03201*.

[20] J. Cho and J. Park, "Multi-modal transformer for brain tumor segmentation," in Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, S. Bakas et al., Eds. Cham, Switzerland: Springer, 2023, pp. 138–148.

[21] R. A. Zeineldin, M. E. Karar, O. Burgert, and F. Mathis-Ullrich, "Multimodal CNN networks for brain tumor segmentation in MRI:

A brats 2022 challenge solution," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, S. Bakas et al., Eds. Cham, Switzerland: Springer, 2023, pp. 127–137.